Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Immunology

An evolutionary perspective of how infection drives human genome diversity: the case of malaria

Valentina D Mangano[1,2] and David Modiano[1,2]

Infection with malaria parasites has imposed a strong selective pressure on the human genome, promoting the convergent evolution of a diverse range of genetic adaptations, many of which are harboured by the red blood cell, which hosts the pathogenic stage of the Plasmodium life cycle. Recent genome-wide and multi-centre association studies of severe malaria have consistently identified *ATP2B4*, encoding the major Ca^{2+} pump of erythrocytes, as a novel resistance locus. Evidence is also accumulating that interaction occurs among resistance loci, the most recent example being negative epistasis among alpha-thalassemia and haptoglobin type 2. Finally, studies on the effect of haemoglobin S and C on parasite transmission to mosquitoes have suggested that protective variants could increase in frequency enhancing parasite fitness.

Addresses
[1] Department of Public Health and Infectious Diseases, University of Rome 'La Sapienza', Rome, Italy
[2] Istituto Pasteur, Fondazione Cenci Bolognetti, University of Rome 'La Sapienza', Rome, Italy

Corresponding authors: Mangano, Valentina D (valentina.mangano@uniroma1.it) and Modiano, David (david.modiano@uniroma1.it)

Current Opinion in Cell Biology 2014, **30**:39–47

This review comes from a themed issue on **Immunogenetics and transplantation**

Edited by **Luis B Barreiro** and **Lluis Quintana-Murci**

For a complete overview see the Issue and the Editorial

Available online 1st July 2014

http://dx.doi.org/10.1016/j.coi.2014.06.004

Introduction: selection from malaria in humans

It is apparent that malaria transmission in humans has become stable with the advent of agriculture around 10,000 years ago, when settlements led to a favourable increase in both human and vector populations density [1]. In 2012, the World Health Organisation reported 207 million cases of malaria (Box 1) and 627,000 deaths worldwide [2]. Malaria incidence and mortality have shown a decrease over the last century in Africa, where most of the burden occurs [3•]: it is thus fair to assume that the historical global burden of malaria has been at least as large as that currently measured. Malaria is therefore considered a strong selective factor on the *Homo sapiens* genome [4], as it has exerted a (relatively) long and large impact on human fitness.

It has been proposed in the 1940–50 that the paradoxically high frequency of deleterious erythrocyte mutations such as alpha-thalassemia in the Mediterranean region [5,6], or sickle cell disease in African populations [7,8], could result from a survival advantage of heterozygotes against malaria. A large body of epidemiological evidence has since then accumulated in favour of the 'malaria hypothesis', showing association of inherited red blood cell disorders with protection against severe malaria (Table 1). Moreover, molecular signatures of selection, such as extended haplotype homozygosity or high Fst [9], have been observed at loci bearing malaria resistant alleles, for instance at *DARC* [10], *G6PD* [11–17], *CD40L* [12] and *CD36* [18].

This review aims at describing how malaria has driven the evolution of the human genome and its diversity, using red blood cell polymorphisms as a striking example, and describing new insights into the genetic architecture of susceptibility to malaria gained from genome-wide and multi-centre studies.

Parallel adaptations to malaria

Epidemiological studies have provided an impressive amount of data illustrating the convergent evolution of erythrocytes malaria resistance factors (Table 1). The red blood cell, which hosts the parasite life cycle stage responsible for pathogenesis, can indeed be seen as a *matryoshka doll* of genetic adaptations to malaria (Figure 1). Within the same cell, many genes, encoding molecules with different functions and localised in different compartments, harbour protective alleles, from loci encoding haemoglobin α and β chains (*HBA* and *HBB*), to those encoding cell surface antigens (e.g. *ABO*) or enzymes with a key role in the cell metabolism (e.g. *G6PD*). Furthermore, within the same gene different alleles can confer protection, for example structural mutant haemoglobins S and C at the *HBB* locus, or multiple *G6PD* mutations causing enzyme deficiency worldwide [19]. Finally, the same protective allele may have arisen independently multiple times: the S allele has been described on five distinct haplotypes in populations from different geographic areas (reviewed in [6]). Such diversified range of adaptations is likely the result of both the strength of the

Box 1 Human malaria parasites.

P. falciparum

Distributed in tropical and subtropical regions of the world, is the most prevalent species (up to 90% of malaria cases) in Sub-Saharan Africa. Responsible of severe malaria syndromes of severe malaria anaemia, cerebral malaria and respiratory distress, is the leading cause of malaria deaths at the global level. Parasites are able to invade erythrocyte of all ages.

P. vivax

The most prevalent species (up to 90% of cases) in Asia and South-America. The low prevalence in Sub-Saharan Africa (1–10%) is thought to be because of the absence of the Duffy antigen, the erythrocyte receptor for the vivax protein mainly responsible for cell invasion, in the human population, although recent studies reported cases in Duffy negative individuals. Previously thought to be a benign form of malaria, increasing evidence shows that *P. vivax* can cause severe symptoms. Parasites can be dormant in the liver for up to 2 years and cause relapses. Reticulocytes only can be invaded.

P. malariae* and *P. ovale

These two species represent up to 3% and 8% of cases, respectively, in Sub-Saharan Africa, and sporadic cases are detected in other tropical and subtropical regions of the world. Their epidemiology is little studied and is likely that the clinical burden is underestimated, although it is apparent that both species cause mild symptoms only. *P. ovale* can be dormant in the liver for up to 4 years.

P. knowlesi

This species primarily infect macaques, but is becoming more prevalent in humans, with reports from South-east Asia indicating that it can represent up to 70% of the malaria cases in some areas.

selective pressure imposed by the Plasmodium parasite, and its complex and plastic interaction with the host cell metabolism and structure [20].

Interaction among malaria protective alleles

Evolutionary interaction among protective alleles should be expected, given their co-occurrence in the same population and individuals, and the related function of the proteins encoded. At the population level, the change in frequency of a protective variant resulting from a new mutation event, or from admixture with a different population, will depend on the pre-existence of protective alleles and their frequency. At the individual level, fitness will depend on the combination of protective variants possessed, and thus selection of distinct alleles will not be independent of each other (selective interaction). This holds true whether or not positive or negative epistasis also occurs (biological interaction).

Selective interaction is exemplified by haemoglobin S and C variants, which co-occur in many populations from West-Africa [21,22]. Epidemiological data indicate that the C allele protects against severe malaria in an additive way, with homozygotes showing higher protection than heterozygotes, and with not associated cost [22–24]. The S allele is instead protective in the heterozygote state, while homozygotes suffer from severe sickle-cell anaemia, an early lethal condition in absence of appropriate medical care [25•]. Population genetic analysis based on relative fitness levels of the six haemoglobin genotypes suggests that the C allele would eventually reach fixation [26], but at a rate dependent on the frequency of the S allele [6].

Negative epistasis between haemoglobin S and α-thalassemia has been demonstrated, showing that the protection afforded by each condition inherited alone is lost when the two conditions are inherited together [24,27]. It has been suggested that such negative epistatic interaction could explain the complex distribution of these haemoglobinopathies in Sub-Saharan African [27], Mediterranean [28] and South-east Asian [29] populations. Similarly, a very recent investigation showed that α-thalassemia modifies the effect of haptoglobin variants on severe malaria: while the Hp2-2 genotype confers a higher risk of disease in children with α-thalassemia, no effect can be seen in children with wild-type haemoglobin. Previous contrasting findings on the role of haptoglobin genotypes in severe malaria could thus be the result of different frequencies of α-thalassemia in the populations under study [30••].

What allelic theory for malaria susceptibility?

Three main theories have been proposed to describe the genetic architecture of susceptibility to disease [31•]. Briefly, the common disease common variant theory [32], built about non-communicable diseases and upon which genome-wide association studies (GWAS) have been designed, hypothesises that common (relatively high frequency) genetic variants account for most of the heritable component of common diseases. The field of primary immunodeficiencies has provided support to the hypothesis that very rare variants with high penetrance could be responsible for disease genetic susceptibility [33,34]. Finally, it has been proposed that for many infectious diseases, heritability could be explained by the cumulative effect of many rare variants with modest penetrance. That this could be the case of malaria has been suggested by a benchmark study that investigated the relative contribution of genetic and non-genetic factors in susceptibility to malaria using pedigree-based variance component analysis. It was estimated that 25% of the total phenotypic variation was explained by additively acting host genes and that haemoglobin S contributed by only 2%, providing evidence for the existence of many unknown protective genes, each individually resulting in small population effects [35]. Under this scenario, GWAS could be not sufficiently powered to uncover the responsible variants, as they are based on